STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:3035–3054 (DOI: 10.1002/sim.1241)

Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random

John S. Preisser[1,*,†], Kurt K. Lohman[2] and Paul J. Rathouz[3]

[1] *Department of Biostatistics, CB #7420, School of Public Health, University of North Carolina, Chapel Hill, NC 27599, U.S.A.*
[2] *Section on Biostatistics, Department of Public Health Sciences, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, U.S.A.*
[3] *Department of Health Studies, University of Chicago, 5841 S. Maryland Ave., MC 2007, Chicago, IL 60637, U.S.A.*

SUMMARY

The generalized estimating equations (GEE) approach is commonly used to model incomplete longitudinal binary data. When drop-outs are missing at random through dependence on observed responses (MAR), GEE may give biased parameter estimates in the model for the marginal means. A weighted estimating equations approach gives consistent estimation under MAR when the drop-out mechanism is correctly specified. In this approach, observations or person-visits are weighted inversely proportional to their probability of being observed. Using a simulation study, we compare the performance of unweighted and weighted GEE in models for time-specific means of a repeated binary response with MAR drop-outs. Weighted GEE resulted in smaller finite sample bias than GEE. However, when the drop-out model was misspecified, weighted GEE sometimes performed worse than GEE. Weighted GEE with observation-level weights gave more efficient estimates than a weighted GEE procedure with cluster-level weights. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: correlated data, drop-outs, estimating equations, logistic models, repeated measures

1. INTRODUCTION

In longitudinal studies, subjects often have data missing due to missed visits, commonly when subjects drop-out of a study or are lost to follow-up. A subject is called a drop-out when the response variable is observed through a certain visit and is missing for all subsequent visits [1]. The problem of drop-outs can be particularly acute in epidemiological cohort studies where interest lies in estimating trends over time and where subjects are followed prospectively over

* Correspondence to: John S. Preisser, Department of Biostatistics, CB #7420, School of Public Health, University of North Carolina, Chapel Hill, NC 27599, U.S.A.
† E-mail: jpreisse@bios.unc.edu

Contract/grant sponsor: American Cancer Society; contract/grant number: IRG-198A
Contract/grant sponsor: National Institutes of Health; contract/grant number: AG14131

Received August 1999
Accepted January 2002

a period spanning several years. The GEE approach [2] has been widely applied in estimating time trends from incomplete longitudinal binary data such as those arising in cohort studies. The performance of GEE for unequal cluster sizes and binary outcomes has been evaluated for between-cluster effects such as found in geographically clustered data [3] and in dose–response studies in toxicology [4] and teratology [5]. Other simulation studies restrict focus to equal cluster sizes [6–9]. Lipsitz *et al.* [10] evaluate GEE by simulation for a bivariate binary response when drop-out occurs prior to the second time point. Evaluation by simulation of the GEE method for estimating time trends from repeated binary data with more than two time points in the presence of drop-outs appears missing from the literature.

To address this deficiency, this paper presents a simulation experiment motivated, in part, by an analysis of cigarette smoking trends among young adults in the Coronary Artery Risk Development in Young Adults (CARDIA) study [11]. In that study, more than 5000 black and white males and females were followed for 7 years and the outcome, self-reported cigarette smoking status (yes/no), was recorded on four occasions. About two-thirds of the participants had complete data. The probability of drop-out conditional upon not dropping out at earlier visits was greater for those who were smokers at previous visits than for those who were non-smokers. In this case, an analysis based upon GEE may give biased estimates of parameters for the regression model of the marginal probabilities describing trends in smoking prevalence. Given the limitations of GEE, there is a growing interest in weighted generalized estimating equations [11–16], which give consistent estimates when the model for the missing data has been correctly specified and consistently estimated.

The validity of an estimating procedure for regression model parameters depends upon model assumptions, including those for missing data. In longitudinal models, the drop-out mechanism may be described hierarchically as data missing completely at random (MCAR), data missing at random, or data missing not at random (MNAR) [17]. The data are MCAR if the drop-out and measurement processes are independent. Under this scenario, the probability of missingness for any response depends in no way on the responses or covariates, whether observed or unobserved. The data are missing at random if missingness depends on observed data but not on unobserved data. In repeated measures studies with drop-outs, and where interest lies in a regression model, two types of missing at random should be delineated. First, covariate dependent missingness allows the missingness of responses to depend only upon covariates. Second, an assumption of missing at random depending upon observed outcomes, which we denote MAR, allows missingness to depend on the observed responses from previous visits as well as observed covariates. Conditional upon these observables, missingness does not depend upon the current or future responses. Finally, missing data is non-ignorable or missing not at random (MNAR) if the missingness is related to the unobserved responses [18]. Neither GEE nor weighted GEE, as considered here, is appropriate in this case. Whether the missingness process is MAR, as opposed to MNAR, often involves assumptions which cannot be tested with the data. Weighted GEEs valid under MNAR [19] are not considered in this paper.

We report on simulation study results that evaluate the performance of GEE and weighted GEE under possible misspecification of models for drop-outs and intraperson correlation. We evaluate bias, test size, coverage of nominal 95 per cent confidence intervals, and relative efficiency under different working correlation structures. Two different formulations of weighted GEE, due to Robins *et al.* [12] and to Fitzmaurice *et al.* [13], are studied with data simulated under several different missingness mechanisms.

2. WEIGHTED ESTIMATING EQUATIONS FOR LONGITUDINAL DATA

We consider marginal models for longitudinal studies that relate the expected value of an individual's binary response at time t (for example, the probability of being a smoker at time t) to covariates via a known link function $g(\cdot)$ [20]. Assume a set of T observation times common to all individuals, and define the complete data response vector $Y_i = (Y_{i1}, \ldots, Y_{iT})'$ for individual i, $i = 1, \ldots, K$. For example, if a subject possesses the outcome of interest (subject is a smoker) $Y_{it} = 1$, and $Y_{it} = 0$ if the subject does not possess this characteristic (subject is not a smoker). The corresponding complete covariate matrix for individual i is $X_i = (X_{i1}, \ldots, X_{iT})'$ where X_{it} is a $p \times 1$ covariate vector at time t.

Now suppose individual i is observed at times $t = 1, \ldots, T_i$, giving the $T_i \times 1$ vector of observed responses, $Y_i^o = (Y_{i1}, \ldots, Y_{iT_i})'$, $1 \leqslant T_i \leqslant T$. Likewise, let $X_i^o = (X_{i1}, \ldots, X_{iT_i})'$ represent the corresponding matrix of covariates. Next, let $E(Y_{it} \mid X_i) = \pi_{it}$ be the probability that $Y_{it} = 1$ given X_i, and let $\pi_i = (\pi_{i1}, \ldots, \pi_{iT})'$ and $\pi_i^o = (\pi_{i1}, \ldots, \pi_{iT_i})'$. Our goal is to make inferences about marginal probabilities, $\pi_{it}(\beta)$, where $g(\pi_{it}) = X_{it}'\beta$, for individuals regardless of whether or not they drop out. In particular, our interest is in generalized linear models for π_{it} given by the logit link function, $g(\pi_{it}) = \ln[\pi_{it}/(1 - \pi_{it})]$.

To estimate β, while accounting for correlation among an individual's repeated responses, Liang and Zeger [2] proposed generalized estimating equations

$$\sum_{i=1}^{K} D_i^{o'}(X_i^o, \beta)(V_i^o)^{-1}[Y_i^o - \pi_i^o(\beta)] = 0 \tag{1}$$

where $D_i^o(X_i^o, \beta) = \partial \pi_i^o / \partial \beta$ is a $T_i \times p$ matrix and V_i^o is a working covariance matrix for Y_i^o. A working model for the correlation is assumed such that $V_i^o = A_i^o C_i^o A_i^o$, $A_i^o = \operatorname{diag}\{v_{it}^{1/2}\}$ is a $T_i \times T_i$ diagonal matrix where $v_{it} = \pi_{it}(1 - \pi_{it})$ for binary data, and $C_i^o(\rho)$ is a working correlation matrix that depends on an unknown nuisance parameter vector ρ. Equation (1) yields a consistent estimate of β if the data are MCAR or missingness depending only on the vector of covariates, whereas, under MAR, a GEE analysis may give biased estimates.

An alternative estimating equations approach that can provide unbiased inference in longitudinal studies with drop-outs has been proposed by Robins *et al.* [12]. They proposed a class of estimating equations in which observations or person-visits have weights inversely proportional to their probability of being observed. This weighted generalized estimating equations approach, which has been called the inverse probability of censoring weighted GEE estimator, is valid under an MAR assumption even if the correlation model is misspecified, provided the model for estimating the probability for missing response is correctly specified. A consistent estimate of β may be obtained from

$$\sum_{i=1}^{K} D_i'(X_i, \beta) V_i^{-1} W_i [Y_i - \pi_i(\beta)] = 0 \tag{2}$$

where $D_i(X_i, \beta) = \partial \pi_i / \partial \beta$ and $V_i = A_i C_i A_i$ is a $T \times T$ working covariate matrix for Y_i. Here, the drop-out process is taken into account through specification of a $T \times T$ diagonal matrix of occasion-specific weights, $W_i = \operatorname{diag}\{R_{i1} w_{i1}, \ldots, R_{iT} w_{iT}\}$, where $R_{it} = 1$ if the ith individual's response is observed at time t, and 0 otherwise. In other words, the weight is given by w_{it} for observed visits and 0 for unobserved visits. The definition and estimation of w_{it}, the inverse of the probability that the ith individual is observed at the tth visit, is described below. Note that unless $C_i = I$, (2) depends upon covariates, assumed known, from both observed and unobserved occasions. The choice of the working correlation matrix in (2) affects efficiency [14]. Like GEE, the use of sandwich variance estimators in weighted GEE provides robustness, in an asymptotic sense, to misspecification of the correlation structure. With consistent estimation of weights provided by a correctly specified drop-out model, weighted GEE does not require correct specification of the correlation structure in order to estimate β and the variance of its estimate consistently.

To characterize the missing data process and obtain estimates $\hat{w}_{it}$, let $\lambda_{it} = P(R_{it} = 1 \mid R_{i(t-1)} = 1, X_i, Y_i, \alpha)$ denote the probability of observing a response at time t for the ith individual conditional on the individual being observed at the time $t-1$. Following the convention of Robins *et al.* [12], we do not consider intermittent missing data patterns, that is, $(R_{it} = 0, R_{i(t+k)} = 1$, for $k > 0$ is not allowed). For the first time point, assume $R_{i1} = 1$ and define $\lambda_{i1} = 1$. The MAR assumption of the weighted GEE approach specifies that $\lambda_{it} = P(R_{it} = 1 \mid R_{i(t-1)} = 1, X_i, (Y_{i1}, \ldots, Y_{i(t-1)}), \alpha)$, so that conditional upon the past responses $Y_{i1}, \ldots, Y_{i(t-1)}$, R_{it} and Y_{it} are independent. The missing data mechanism therefore depends only on observed data and may be specified up to a $q \times 1$ vector of unknown parameters, α. We obtain $\hat{\lambda}_{it}$ by fitting a logistic model, $\mathrm{logit}\{\lambda_{it}(\alpha)\} = Z_{it}\alpha$, with a vector of predictors, Z_{it} which may include visit indicator variables, covariates and past responses. The log partial likelihood for the ith subject is

$$\sum_t R_{i,t-1} \log\{\lambda_{it}(\alpha)^{R_{it}}[1 - \lambda_{it}(\alpha)]^{1-R_{it}}\} \tag{3}$$

Differentiation with respect to α yields $S_i(\alpha) = \sum_t R_{i,t-1} Z_{it}[R_{it} - \lambda_{it}(\alpha)]$, the score component of the ith individual. The partial likelihood estimate $\hat{\alpha}$ and thus $\hat{\lambda}_{it}$'s are obtained by solving the estimating equations $\sum_i S_i(\alpha) = 0$. The weight w_{it} for the ith individual at the tth time is the inverse of the unconditional probability of being observed at time t, estimated as the inverse of the cumulative product of conditional probabilities, $\hat{w}_{it}^{-1} = \hat{\lambda}_{i1} \times \cdots \times \hat{\lambda}_{it}$. Note that an observation with a low probability of being observed will receive a large weight.

The weighted GEE estimate is obtained by solving (2) where $W_i(\hat{\alpha})$ is a diagonal matrix with elements $\hat{w}_{it}$, for $t = 2, \ldots, T_i$, and $\hat{w}_{i1} = 1$. As in GEE, iteratively reweighted estimation of β alternates at each step with method of moments estimation of ρ. Under correctly specified models for the marginal means and for the drop-out process, equation (2) yields a consistent estimate of β which has an asymptotic normal distribution with consistent estimator of its asymptotic variance given by

$$\left(\sum_{i=1}^{K} D_i' V_i^{-1} W_i D_i\right)^{-T} \sum_{i=1}^{K} E_i E_i' \left(\sum_{i=1}^{K} D_i' V_i^{-1} W_i D_i\right)^{-1} \tag{4}$$

where $E_i = U_i - (\sum_{i=1}^{K} U_i S_i')(\sum_{i=1}^{k} S_i S_i')S_i$, $U_i = D_i' V_i^{-1} W_i (Y_i - \pi_i)$, and S_i is the score component for the ith individual from the drop-out model [12]. Note that $\sum_{i=1}^{K} E_i E_i'$ is the matrix sum of squares and cross-products of the residuals from the multivariate regression of the score component from (2) on the score vectors from the drop-out model (3). The use of $\sum_{i=1}^{K} E_i E_i'$ in the centre of (4) instead of $\sum_{i=1}^{K} U_i U_i'$ adjusts for estimation of α.

Fitzmaurice *et al.* [13] studied a weighted GEE using cluster level weights. Their weighted estimating equations are

$$\sum_{i=1}^{K} w_i D_i^{o'}(X_i^o,\beta)(V_i^o)^{-1}[Y_i^o - \pi_i^o(\beta)] = 0 \tag{5}$$

with cluster-level (subject-specific) weights, w_i. An individual's weight is the inverse of the probability of dropping out at the observed time of drop-out. Letting m denote the time of drop-out, $2 \leqslant m \leqslant T+1$, for the ith individual

$$w_i^{-1} = \left(\prod_{t=2}^{m-1} \lambda_{it}\right)(1-\lambda_{im})^{I\{m\leqslant T\}}$$

where $I(\cdot)$ is the indicator function. Again, following Robins *et al.* [12], a consistent estimator for the variance of $\hat{\beta}$ is

$$\left(\sum_{i=1}^{K} w_i D_i^{o'}(V_i^o)^{-1} D_i^o\right)^{-T} \sum_{i=1}^{K} E_i^o E_i^{o'} \left(\sum_{i=1}^{K} w_i D_i^{o'}(V_i^o)^{-1} D_i^o\right)^{-1} \tag{6}$$

where $E_i^o = U_i^o - (\sum_{i=1}^{K} U_i^o S_i')(\sum_{i=1}^{k} S_i S_i')S_i$ and $U_i^o = w_i D_i^{o'}(V_i^o)^{-1}(Y_i^o - \pi_i^o)$.

3. A SIMULATION STUDY

3.1. Design

A simulation experiment was performed to compare six methods of analysis under different conditions pertaining to the correlation structure, the missing data mechanism and the amount of missingness. The methods include GEE, given by (1), and the two weighted GEE methods, (2) and (5), respectively, each under independence and exchangeable working correlation structures. The design of the simulation study was based partly upon characteristics of data from the CARDIA study where the binary response was self-described smoking status at time t. For each individual, a vector of correlated binary responses, $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})$, indicating smoking status at T time points were generated, where the marginal log-odds of being a smoker at time $t = 1, 2, \ldots, T$ was taken to be $\text{logit}[P(Y_{it}=1)] = \beta_1 + \beta_T(\frac{t-1}{T-1})$. Throughout, we fixed $\beta_1 = -0.7$ and $\beta_T = 0.2$ corresponding to marginal probabilities $\pi_{i1} = 0.332$ and $\pi_{iT} = 0.378$ indicating a moderate increase in smoking prevalence. Correlated binary responses were generated using a method based upon a family of multivariate binary distributions with a certain conditional linear property. This method requires specification of the $T \times 1$ vector of marginal means π_i, and the $T \times T$ correlation matrix, C_i. See the Appendix for a detailed description of the data generating process.

Data were generated under three correlation structures: independence given by $C_i = I$, and exchangeable correlation structures with the common correlation among any two time points taking values of $\rho = 0.2$ and $\rho = 0.6$. These three correlation structures represent no correlation, weak correlation and strong correlation, respectively. The strong correlation case resembles data from the CARDIA study where pairwise correlations are high irrespective of the time between measures because there were relatively few young adults who initiated or quit smoking over the course of the study. Each method was applied with $K=50$, $K=100$ and $K=200$ subjects. We set $T=6$ for the first two cases, and $T=4$ when $K=200$.

Observed data, Y_i^o, were generated according to simulated patterns of missingness given by the indicators $R_{i2},\dots,R_{iT}$ generated under various models for the drop-out process. We assumed that $R_{i1}=1$ with probability 1 and set $R_{i,t+k}=0, k>0$ whenever $R_{it}=0$ so that intermittent missing data patterns are not allowed. Missingness models are of the form

$$\text{logit}(\lambda_{it})=\alpha_0+\alpha_1 y^\star_{i(t-1)}+\alpha_2 y^\star_{i(t-2)}I(t>2)+\alpha_3 y^\star_{it} \quad t=2,\dots,T \tag{7}$$

where $I(t>2)=1$ if $t>2$, and 0 otherwise, and $y^\star_{it}=2y_{it}-1$ (y_{it} is a realization of the random variable Y_{it}) giving $y^\star_{it}=1$ if the ith individual was a smoker at time t and -1 if a non-smoker. Model (7) specifies that λ_{it}, the probability of being observed at time t, given being observed at time $t-1$, may depend on the smoking status at the current or previous two observations. Indicators for non-missingness, $R_{i2},\dots,R_{iT}$, were generated from five different general models for drop-out determined by $\alpha=(\alpha_0,\alpha_1,\alpha_2,\alpha_3)$. The five missing data mechanisms are given by (i) $\alpha=(\alpha_0,0,0,0)$, (ii) $\alpha=(\alpha_0,-0.2,0,0)$, (iii) $\alpha=(\alpha_0,-0.5,0,0)$, (iv) $\alpha=(\alpha_0,-0.5,-0.2,0)$, and (v) $\alpha=(\alpha_0,0,0,-0.5)$. Case (i) is an MCAR process, cases (ii) and (iii) are MAR (weak and strong, respectively), case (iv) is a more complicated MAR process (which we call two-dependent) where drop-out depends upon smoking status at the two previous time points. Finally, case (v) is MNAR since drop-out depends upon the potentially unobserved value of smoking at the current time point. These simulations were repeated by considering four different levels of α_0, a parameter that roughly speaking relates to the average conditional probability of drop-out under any given model (7). We consider values of 3.0, 2.2, 1.4 and 0.4 for α_0. These four values specify the 'average' probability of drop-out at the current visit given not having dropped out prior to the visit to be 0.05 ('minimal'), 0.10 ('mild'), 0.20 ('moderate') and 0.40 ('severe'), respectively. For $K=50$ we do not report severe drop-out since resulting data contains too little information to justify use of the methods considered here, and for $K=100$ or $K=200$ we do not consider minimal drop-out since preliminary investigation found the results were similar to the case of mild drop-out.

For each of the 45 ways that data were generated (three correlation values $\times$ five general missing data models $\times$ three overall magnitudes of missingness per scenario), we compared the six methods of analysis using 1000 replicate observations. For weighted GEE, we used an estimated missing data model like (7) but that included only an intercept and smoking status at the previous visit (that is, $\alpha_2=\alpha_3=0$). Thus, the first three cases of the missingness generating process above would expect to yield approximately unbiased weighted GEE estimates of β, while the two-dependent MAR and the MNAR process would correspond to misspecified drop-out models. The parameter of interest is β_T, the change in the log-odds of smoking from the last compared to the first time point. We fitted the unconstrained (if overparameterized) marginal log-odds model, $\text{logit}[P(Y_{it}=1)]=\beta_1+\beta_t I\ (t>1),\ t=1,\dots,T$, estimating T regression parameters instead of 2. None the less, β_T in the fitted model retains the same interpretation as in the model which generated the data. We specified binomial-type variances, $\text{var}(Y_{it})=\pi_{it}(1-\pi_{it})$, and used Liang and Zeger's [2] all-available-pairs estimator to estimate ρ.

For each of the six methods of analysis, we evaluate, for β_T, per cent relative bias, test size, observed coverage of a nominal 95 per cent confidence interval, and the accuracy of its variance estimator. Per cent relative bias was computed as $(1/1000)\sum_{s=1}^{1000}(\hat{\beta}_s-\beta_T)/\beta_T\times 100$

per cent where $\beta_T = 0.2$ and $\hat{\beta}_s$ is the estimate of β_T from the sth simulated replicate. Test size was defined as the proportion of times $|\hat{\beta}_T/\mathrm{SE}(\hat{\beta}_T)| \geqslant t_{0.975,K-T}$ when $\beta_T = 0$, where $t_{0.975,K-T}$ is the 97.5th percentile of Student's t-distribution with $K - T$ degrees of freedom, and 'SE' denotes the empirical sandwich standard error for GEE or the standard error from (4) or (6) for the respective weighted GEE methods. Coverage was defined as the per cent of 95 per cent Wald-type confidence intervals for β_T using the same standard error formulae as for size that contained $\beta_T = 0.2$. Finally, the standard errors for the six methods are evaluated by comparing their average values over all simulations to the Monte Carlo or empirical standard deviation of the 1000 $\hat{\beta}_T$ parameter estimates.

In addition to examining the four criteria described above, the efficiency of the GEE and weighted GEE estimators relative to a maximum likelihood estimator was determined for β_4 when $K = 200$. The relative efficiency was defined as the ratio of the Monte Carlo mean squared error of the maximum likelihood estimator $\hat{\beta}_{4,\mathrm{ML}}$, to that of estimating equations estimator, $\hat{\beta}_4$. Our rationale for examining a maximum likelihood estimator is that the weighted GEE estimators given by (2) and (5) are not in general the semi-parametric efficient estimators in their classes. We consider the maximum likelihood estimator $\hat{\beta}_{4,\mathrm{ML}}$ based upon the unconstrained multinomial likelihood model for the 2^4 contingency table formed by the cross-classification of the binary smoking outcomes from the four time points. This model places no restrictions on the 16×1 vector of joint cell probabilities, θ, other than that its elements sum to 1. Thus, the 'working' model for maximum likelihood estimation is not the same as the data generation model which it has as a special case. The 'working' model for maximum likelihood coincides with working models used in the GEE and weighted GEE methods with respect to the marginal mean structure. However, unlike the estimating equations methods considered, its implicit working covariance structure is unstructured since no constraints are placed on the joint probabilities. For $T = 4$, $\hat{\beta}_{T,\mathrm{ML}}$ is easily determined as a matrix function of $\hat{\theta}$, the maximum likelihood estimator of θ. In particular, $\hat{\beta}_{4,\mathrm{ML}} = A_3 \log A_2 \hat{\theta}$ with A_2 a matrix of 0's and 1's transforming the joint cell probabilities to the marginal probabilities, the log operator defined elementwise, and A_3 a constrast vector of -1's, 0's and 1's. Finally, under the unconstrained multinomial model, $\hat{\theta}$ has a closed-form expression based upon factored likelihoods for multinomial data with monotone patterns of missingness (Little and Rubin, reference [17], Section 9.2). When there is complete data (that is, $T_i = T$ for $i = 1, \ldots, K$), $\hat{\theta}$ is simply the vector of observed cell relative frequencies.

3.2. *Results*

The results of the simulation study are given in the tables where 'WEE' refers to the observation-weighted GEE procedure using (2) and 'CWEE' refers to the cluster-weighted GEE procedure using (5). Results for the WEE method with independent working correlation matrix (WEE-indep) are not shown in the tables since these results were nearly the same as the results for WEE with exchangeable working correlation (WEE-exch). We comment with a footnote in a table whenever WEE-indep gives a different result than WEE-exch.

Tables I to III report on bias of $\hat{\beta}_T$ for $K = 50$, $K = 100$ and $K = 200$, respectively. In each table, all methods have large bias under MNAR as expected. We consider per cent relative bias to be low if it is at most 25 per cent, and to be large otherwise. Then, under those cases where the asymptotic bias is zero, that is all methods under MCAR missingness,

Table I. Per cent relative bias of $\hat{\beta}_6$, $K=50$, $T=6$.

ρ	True model		Working model				
	Drop-out		GEE		WEE	CWEE	
	Type	Amount	Indep	Exch	Exch	Indep	Exch
0.0	MCAR	5%	−8	−8	−8	26	30
	MAR-weak		−4	−4	−4	15	7
	MAR-strong		7	6	7	40	13
	MAR-2-dep		9	10	9	12	−14
	MNAR		−21	−21	−21	26	27
0.2	MCAR		6	5	6	22	18
	MAR-weak		−5	4	5	42	23
	MAR-strong		−32	−9	−9	15	−2
	MAR-2-dep		−29	0	−4	−19	−19
	MNAR		−43	−33	−38	−80	−75
0.6	MCAR		0	0	−0	17	−0
	MAR-weak		−29	−3	−2	19	3
	MAR-strong		−70	2	5	8	3
	MAR-2-dep		−97	−3	−12	−49	−11
	MNAR		−83	−29	−41	−90	−62
0.0	MCAR	10%	−10	−10	−10	1	−5
	MAR-weak		−7	−8	−7	3	−7
	MAR-strong		7	6	8	16	−2
	MAR-2-dep		6	6	5	−7	−35
	MNAR		−55	−55	−55	−38	−35
0.2	MCAR		−1	−2	−1	1	2
	MAR-weak		−17	2	3	14	10
	MAR-strong		−61	−13	−12	−6	−12
	MAR-2-dep		−64	−3	−14	−38	−25
	MNAR		−94	−72	−83	−123	−108
0.6	MCAR		−3	−5	−5	−2	−5
	MAR-weak		−64	−5	−3	4	2
	MAR-strong		−149	−5	1	5	3
	MAR-2-dep		−203	−16	−34	−59	−12
	MNAR		−175	−63	−88	−136	−87
0.0	MCAR	20%	−30	−31	−29	−28	−32
	MAR-weak		−30	−30	−31	−31	−39
	MAR-strong		3	1	5	9	−5
	MAR-2-dep		−9	−8	−12	−2	−23
	MNAR		−110	−109	−111	−107	−107
0.2	MCAR		−12	−15	−13	−11	−13
	MAR-weak		−49	−8	−9	−8	−5
	MAR-strong		−124	−32	−33	−25	−26
	MAR-2-dep		−129	−17	−41	−39	−7
	MNAR		−191	−154	−174	−191	−171
0.6	MCAR		−9	−9	−10	−12	−7
	MAR-weak		−136	−15	−17	−13	−3
	MAR-strong		−299	−35	−22	−17	−2
	MAR-2-dep		−392	−53	−110	−116	−18
	MNAR		−349	−141	−195	−215	−133

WEE-indep per cent relative bias is the same as the reported WEE-exch per cent relative bias rounded to nearest whole per cent.

Table II. Per cent relative bias of $\hat{\beta}_6$, $K=100$, $T=6$.

ρ	True model		Working model				
	Drop-out		GEE		WEE	CWEE	
	Type	Amount	Indep	Exch	Exch	Indep	Exch
0.0	MCAR	10%	−12	−12	−12	−9	−12
	MAR-weak		−4	−4	−4	1	−4
	MAR-strong		−4	−4	−3	−3	−12
	MAR-2-dep		1	2	2	−13	−34
	MNAR		−47	−47	−47	−34	−30
0.2	MCAR		−2	−3	−2	0	−2
	MAR-weak		−20	−1	1	5	2
	MAR-strong		−55	−6	−4	2	0
	MAR-2-dep		−67	−7	−17	−40	−20
	MNAR		−97	−77	−86	−124	−107
0.6	MCAR		−1	−3	−2	1	−1
	MAR-weak		−60	−3	−0	2	−1
	MAR-strong		−150	−8	−2	−1	−0
	MAR-2-dep		−199	−14	−30	−59	−9
	MNAR		−172	−61	−85	−134	−84
0.0	MCAR	20%	−24	−25	−24	−24	−26
	MAR-weak		−17	−18	−18	−17	−19
	MAR-strong		−2	−2	−0	0	−5
	MAR-2-dep		−3	−2	−3	3	−10
	MNAR		−97	−97	−97	−94	−94
0.2	MCAR		−7	−10	−7	−6	−8
	MAR-weak		−43	−3	−1	−0	0
	MAR-strong		−107	−15	−12	−9	−8
	MAR-2-dep		−121	−12	−31	−31	1
	MNAR		−182	−148	−164	−181	−161
0.6	MCAR		−7	−7	−7	−9	−6
	MAR-weak		−123	−8	−3	−0	−0
	MAR-strong		−294	−31	−12	−10	−2
	MAR-2-dep		−369	−40	−77	−88	−4
	MNAR		−345	−133	−183	−206	−128
0.0	MCAR	40%	21	15	22	26	21
	MAR-weak		−7	−11	−13	−10	−19
	MAR-strong		−4	−9	−10	−4	−20
	MAR-2-dep		−12	−11	−17	20	−4
	MNAR		−187	−191	−187	−185	−193
0.2	MCAR		−7	−31	−9	−5	−24
	MAR-weak		−54	−10	16	22	6
	MAR-strong		−188	−57	−37	−30	−37
	MAR-2-dep		−231	−58	−103	−68	−33
	MNAR		−324	−315	−296	−281	−286
0.6	MCAR		6	−46	−9	−9	−40
	MAR-weak		−219	−58	−10	−2	−23
	MAR-strong		−466	−109	−64	−39	−33
	MAR-2-dep		−577	−144	−238	−197	−65
	MNAR		−556	−330	−315	−283	−255

WEE-indep per cent relative bias is the same as the reported WEE-exch per cent relative bias rounded to nearest whole per cent except for three cases for 40 per cent drop-out: (i) ($\rho=0$, MAR-weak) $=-14$; (ii) ($\rho=0.6$, MAR-strong) $=-66$; (iii) ($\rho=0.6$, MNAR) $=-314$.

Table III. Per cent relative bias of $\hat{\beta}_4$, $K=200$, $T=4$.

ρ	True model		Working model					
	Drop-out		GEE		WEE	CWEE		ML
	Type	Amount	Indep	Exch	Exch	Indep	Exch	
0.0	MCAR	10%	0	0	0	1	1	−0
	MAR-weak		0	0	0	1	−2	0
	MAR-strong		2	2	2	6	−4	2
	MAR-2-dep		−2	−2	−2	−33	−52	−2
	MNAR		−47	−47	−47	−42	−43	−46
0.2	MCAR		−7	−7	−7	−4	−4	−7
	MAR-weak		−10	1	2	8	7	2
	MAR-strong		−27	2	2	4	1	3
	MAR-2-dep		−35	−1	−5	−44	−35	1
	MNAR		−69	−59	−63	−97	−87	−58
0.6	MCAR		−1	−0	−1	−3	−3	−0
	MAR-weak		−33	1	2	2	2	2
	MAR-strong		−87	−2	1	0	1	1
	MAR-2-dep		−109	−3	−9	−39	−10	3
	MNAR		−110	−46	−55	−101	−76	−43
0.0	MCAR	20%	−4	−5	−5	−4	−4	−5
	MAR-weak		−2	−2	−2	−5	−6	−3
	MAR-strong		−0	−0	−0	2	−1	0
	MAR-2-dep		−5	−5	−4	−23	−33	−3
	MNAR		−95	−95	−95	−92	−92	−93
0.2	MCAR		−8	−7	−7	−9	−7	−7
	MAR-weak		−22	0	1	1	2	1
	MAR-strong		−56	−2	1	−3	−1	2
	MAR-2-dep		−68	−6	−13	−38	−22	−0
	MNAR		−139	−121	−128	−153	−140	−119
0.6	MCAR		−0	−1	−1	−2	−2	−0
	MAR-weak		−70	−2	−0	−1	−0	1
	MAR-strong		−172	−13	−1	−2	−0	−2
	MAR-2-dep		−212	−19	−25	−46	−6	−2
	MNAR		−220	−99	−114	−148	−115	−89
0.0	MCAR	40%	−15	−15	−15	−15	−15	−11
	MAR-weak		−1	−2	−2	−2	−3	4
	MAR-strong		−11	−10	−13	−10	−14	−1
	MAR-2-dep		−3	−1	−1	11	3	13
	MNAR		−200	−200	−200	−200	−201	−186
0.2	MCAR		−4	−5	−5	−5	−4	−4
	MAR-weak		−53	−10	−8	−9	−9	−9
	MAR-strong		−114	−17	−6	−6	−6	−6
	MAR-2-dep		−129	−20	−30	−21	2	−10
	MNAR		−280	−247	−260	−262	−247	−235
0.6	MCAR		−9	−10	−9	−9	−6	−6
	MAR-weak		−152	−15	−14	−15	−4	−8
	MAR-strong		−331	−48	−11	−10	1	−10
	MAR-2-dep		−399	−66	−77	−76	1	−17
	MNAR		−439	−218	−242	−243	−190	−186

WEE-indep relative bias is the same as the reported WEE-exch relative bias rounded to nearest whole per cent.

and the weighted GEE methods under MAR-weak and MAR-strong, the following general observations can be made: (i) for $K = 50$, the bias is low for minimal or mild drop-out, but not for moderate drop-out; (ii) for $K = 100$, the bias is low for mild and moderate drop-out, but not for severe drop-out; (iii) for $K = 200$, the bias is low for mild, moderate and severe drop-out. Henceforth, and in agreement with simulation results to follow, we refer to mild, moderate and severe drop-out as the limiting missingness situations for $K = 50$, $K = 100$ and $K = 200$, respectively, where these drop-out rates are sufficiently low such that they correspond to acceptable performance of $\hat{\beta}_T$ in this study. The one exception to these remarks is that the bias for CWEE is large for minimal drop-out in Table I. Indeed, the bias more than doubles those figures under limited simulations for 2 per cent drop-out (not shown), underscoring the poor properties of the CWEE method in situations where one would want a method to perform best.

All three tables show that use of GEE may result in biased estimates when drop-out is MAR. In particular, GEE with independence working correlation resulted in heavily biased estimates for non-zero correlation. This bias increased in absolute terms as the true correlation increased or as the strength of the MAR effect grew. Interestingly, GEE with exchangeable working correlation gave estimates with small bias under MAR-weak when the correlation was small ($\rho \leqslant 0.2$). However, under limiting situations with MAR-strong and $\rho = 0.6$ (for example, $K = 100$ with 20 per cent drop-out; $K = 200$ with 40 per cent drop-out) GEE with the correct exchangeable working correlation gave large bias. As noted earlier, WEE with independence or exchangeable correlation gave small bias in these cases under correctly specified MAR (weak or strong) mechanisms. However, under a misspecified missingness model (MAR-2-dep), the results show it is possible to have greater bias with weighted GEE than with GEE.

Tables IV to VI report on size. Table IV shows that the CWEE method can be very anti-conservative when drop-out is minimal. When drop-out did not exceed that of the limiting situations for a given K, WEE performed well. Conversely, in some of those same situations, the test size of GEE was inflated, even when the exchangeable correlation was correctly specified. In particular, for $K = 100$ with 20 per cent drop-out, $\rho = 0.6$, and MAR-strong, the test size of GEE-exch was estimated to be 0.089 versus 0.048 for the WEE methods. For $K = 200$ with 40 per cent drop-out, $\rho = 0.6$, and MAR-strong, the test size of GEE-exch was estimated to be 0.108, more than double the nominal 0.05 level, versus 0.051 for the WEE method. The test size for GEE-indep can be very inflated under MAR. Thus the results for test size parallel those for bias in that they are in general agreement with statements made above about the limiting situations of acceptable performance.

Tables VII and VIII report for these limiting situations, coverage of nominal 95 per cent confidence intervals for β_T and performance of the standard errors of $\hat{\beta}_T$, respectively. We consider only MCAR and MAR corresponding to correctly specified drop-out models for the weighted GEE methods. For the weighted GEE methods, these results indicate coverage near the nominal 95 per cent level. Table VII shows that under MAR-strong and $\rho = 0.6$, GEE may result in undercoverage. Table VIII shows that for the three methods that use an exchangeable working correlation matrix, the standard errors tend to slightly underestimate the true values. For $K = 200$, the underestimation was less when drop-out was less (not shown). Results for the independence working correlation are similar to those shown in Table VIII. Generally, these results along with the confidence interval coverage close to the nominal 95 per cent level indicates that the variance estimators (along with asymptotic normality) are reasonable for the limiting situations.

Table IV. Size of β_6 (probability $\times$ 1000), $K=50$, $T=6$.

ρ	True model		Working model				
	Drop-out		GEE		WEE	CWEE	
	Type	Amount	Indep	Exch	Exch	Indep	Exch
0.0	MCAR	5%	40	40	43	88	93
	MAR-weak		39	39	40	101	99
	MAR-strong		46	46	48	118	114
	MAR-2-dep		39	39	40	141	130
	MNAR		39	39	41	98	105
0.2	MCAR		55	50	58	111	120
	MAR-weak		50	55	54	100	92
	MAR-strong		48	50	49	112	113
	MAR-2-dep		52	55	52	101	106
	MNAR		38	35	39	83	101
0.6	MCAR		41	41	47	80	87
	MAR-weak		48	49	50	63	87
	MAR-strong		56	54	46	80	71
	MAR-2-dep		95	64	52	74	74
	MNAR		63	52	57	94	92
0.0	MCAR	10%	44	45	46	53	54
	MAR-weak		42	40	41	60	53
	MAR-strong		51	51	50	67	66
	MAR-2-dep		43	43	44	63	62
	MNAR		31	31	33	46	44
0.2	MCAR		44	51	50	54	48
	MAR-weak		47	50	49	69	59
	MAR-strong		51	51	56	67	67
	MAR-2-dep		45	60	53	73	75
	MNAR		54	51	48	57	59
0.6	MCAR		39	41	45	55	55
	MAR-weak		50	55	53	47	57
	MAR-strong		72	55	39	67	59
	MAR-2-dep		153	73	42	61	67
	MNAR		102	72	61	81	77
0.0	MCAR	20%	40	43	41	37	37
	MAR-weak		34	37	42	41	41
	MAR-strong		42	41	52	44	43
	MAR-2-dep		34	36	48	54	53
	MNAR		18	17	21	22	15
0.2	MCAR		45	48	53	45	48
	MAR-weak		34	49	40	43	46
	MAR-strong		42	58	48	47	47
	MAR-2-dep		50	70	54	68	76
	MNAR		44	48	45	49	57
0.6	MCAR		29	50	39	41	56
	MAR-weak		30	58	42	39	51
	MAR-strong		78	85	59	66	77
	MAR-2-dep		146	115	47	57	84
	MNAR		92	121	31	48	68

WEE-indep test size is the same as the reported WEE-exch test size.

Table V. Size of β_6 (probability$\times$1000), $K = 100$, $T = 6$.

ρ	True model		Working model				
	Drop-out		GEE		WEE	CWEE	
	Type	Amount	Indep	Exch	Exch	Indep	Exch
0.0	MCAR	10%	44	42	45	50	51
	MAR-weak		45	43	45	43	39
	MAR-strong		52	52	54	68	69
	MAR-2-dep		45	47	51	62	69
	MNAR		44	45	45	47	42
0.2	MCAR		50	54	54	50	44
	MAR-weak		52	59	64	61	60
	MAR-strong		56	52	51	50	55
	MAR-2-dep		56	47	43	67	63
	MNAR		80	77	74	86	86
0.6	MCAR		61	54	58	48	53
	MAR-weak		69	50	45	53	46
	MAR-strong		181	61	49	68	68
	MAR-2-dep		266	74	50	61	57
	MNAR		196	76	85	108	98
0.0	MCAR	20%	37	40	42	41	41
	MAR-weak		47	44	46	42	39
	MAR-strong		54	52	52	57	58
	MAR-2-dep		36	41	46	53	56
	MNAR		40	42	41	42	40
0.2	MCAR		40	39	44	50	42
	MAR-weak		43	47	47	55	51
	MAR-strong		66	47	49	45	54
	MAR-2-dep		82	63	54	51	66
	MNAR		105	93	97	104	102
0.6	MCAR		42	48	44	45	44
	MAR-weak		85	57	49	59	46
	MAR-strong		284	89	48	60	66
	MAR-2-dep		417	106	49	53	72
	MNAR		356	172	107	129	108
0.0	MCAR	40%	18	22	23	24	29
	MAR-weak		14	20	18	20	26
	MAR-strong		23	27	46	50	47
	MAR-2-dep		18	22	33	39	41
	MNAR		8	13	8	11	12
0.2	MCAR		12	48	20	23	51
	MAR-weak		15	65	34	40	73
	MAR-strong		6	70	63	68	82
	MAR-2-dep		13	74	55	55	79
	MNAR		17	90	25	19	67
0.6	MCAR		15	92	24	23	78
	MAR-weak		2	133	41	41	112
	MAR-strong		22	191	77	83	151
	MAR-2-dep		96	254	86	77	194
	MNAR		45	360	17	15	194

WEE-indep test size is the same as the reported WEE-exch test size except table entry is 21 for (drop-out = 40 per cent, $\rho = 0.2$, MCAR).

Table VI. Size of β_4 (probability $\times$ 1000), $K=200$, $T=4$.

ρ	True model		Working model					
	Drop-out		GEE		WEE	CWEE		ML
	Type	Amount	Indep	Exch	Exch	Indep	Exch	
0.0	MCAR	10%	49	50	49	47	54	49
	MAR-weak		35	35	34	40	43	38
	MAR-strong		44	44	47	61	62	47
	MAR-2-dep		46	47	48	58	71	48
	MNAR		68	69	69	61	57	68
0.2	MCAR		51	51	51	59	59	52
	MAR-weak		77	73	75	78	76	73
	MAR-strong		49	44	47	68	67	48
	MAR-2-dep		45	39	39	64	70	38
	MNAR		94	82	88	119	121	83
0.6	MCAR		50	59	62	54	60	58
	MAR-weak		71	45	49	52	51	43
	MAR-strong		160	38	36	58	52	40
	MAR-2-dep		238	60	53	63	58	50
	MNAR		261	100	106	140	119	74
0.0	MCAR	20%	53	51	56	52	51	50
	MAR-weak		41	41	43	38	39	42
	MAR-strong		54	52	50	53	51	48
	MAR-2-dep		56	51	53	54	63	53
	MNAR		98	100	101	96	96	98
0.2	MCAR		42	39	40	44	43	50
	MAR-weak		75	74	74	70	75	68
	MAR-strong		59	46	44	41	40	46
	MAR-2-dep		81	43	43	53	51	43
	MNAR		197	159	166	191	181	149
0.6	MCAR		48	61	57	57	52	60
	MAR-weak		101	50	42	47	45	43
	MAR-strong		362	48	39	33	36	37
	MAR-2-dep		508	72	51	63	47	45
	MNAR		538	218	203	230	220	138
0.0	MCAR	40%	55	56	56	52	54	64
	MAR-weak		42	39	43	47	44	39
	MAR-strong		45	44	45	41	41	37
	MAR-2-dep		45	50	52	61	58	44
	MNAR		161	163	167	169	169	155
0.2	MCAR		40	44	44	41	44	47
	MAR-weak		54	59	55	55	60	58
	MAR-strong		85	57	47	50	42	35
	MAR-2-dep		95	50	55	50	51	46
	MNAR		313	291	284	280	267	216
0.6	MCAR		38	51	42	44	57	41
	MAR-weak		117	65	44	43	43	44
	MAR-strong		512	108	51	45	55	33
	MAR-2-dep		670	158	63	56	82	37
	MNAR		721	437	249	232	241	145

WEE-indep test size is the same as the reported WEE-exch test size for all rows.

Table VII. Coverage of nominal 95 per cent confidence intervals for β_T.

K	T	ρ	True model		Working model				
			Drop-out		GEE		WEE	CWEE	
			Type	Amount	Indep	Exch	Exch	Indep	Exch
50	6	0.0	MCAR	10%	94.8	94.6	94.7	94.4	93.9
			MAR-weak		96.3	96.1	95.9	94.4	94.6
			MAR-strong		96.0	96.2	95.4	93.3	93.4
		0.2	MCAR		96.0	95.7	95.2	95.7	95.6
			MAR-weak		94.9	94.2	94.2	93.1	93.5
			MAR-strong		95.1	94.9	94.6	93.9	93.4
		0.6	MCAR		96.0	95.5	95.6	93.8	94.2
			MAR-weak		93.7	95.1	95.2	95.0	95.3
			MAR-strong		91.8	94.7	96.1	93.9	94.5
100	6	0.0	MCAR	20%	94.9	95.2	94.6	95.3	95.8
			MAR-weak		95.2	95.2	95.1	95.5	95.4
			MAR-strong		94.6	94.6	94.4	94.4	94.7
		0.2	MCAR		94.6	94.7	94.5	94.8	94.5
			MAR-weak		94.8	94.9	94.5	95.4	95.1
			MAR-strong		93.2	94.3	94.3	95.0	94.5
		0.6	MCAR		95.4	94.6	95.0	95.8	95.0
			MAR-weak		92.0	94.5	94.0	93.3	94.5
			MAR-strong		71.3	92.5	95.7	94.9	94.0
200	4	0.0	MCAR	40%	94.3	94.1	93.8	94.1	94.1
			MAR-weak		95.4	95.6	95.4	95.4	95.3
			MAR-strong		96.2	96.2	95.8	94.9	94.8
		0.2	MCAR		95.0	95.0	94.9	95.5	95.4
			MAR-weak		94.0	94.1	93.9	94.1	93.8
			MAR-strong		91.2	93.9	94.5	95.4	95.3
		0.6	MCAR		96.7	95.4	95.5	95.5	95.7
			MAR-weak		88.6	94.0	96.1	96.4	96.2
			MAR-strong		48.7	88.6	95.1	95.1	95.1

WEE-indep coverage is the same as the reported WEE-exch coverage rounded to nearest tenth except for three cases: (i) (drop-out = 20 per cent, $\rho = 0.2$, MAR-strong)= 94.4; (ii) (drop-out = 40 per cent, $\rho = 0$, MAR-weak)= 97.4; (iii) (drop-out = 40 per cent, $\rho = 0.6$, MAR-weak)= 96.0.

Table IX reports per cent relative efficiency with respect to maximum likelihood of GEE and the two weighted GEE methods for $K = 200$, $T = 4$. The efficiency of WEE, like that of GEE with correctly specified exchangeable working correlation, was generally high. Efficiency for these methods decreased with increasing drop-out and correlation as illustrated with 40 per cent drop-out and $\rho = 0.6$ when efficiency of WEE and GEE-exch was 70 per cent and 26 per cent, respectively. The low efficiency for GEE-exch in this case is related to its large bias as shown in Table III. We note that when $T = 4$ under 40 per cent drop-out, only 22 per

Table VIII. Average standard errors (SE) and Monte Carlo standard deviation (SD) of $\hat{\beta}_T(\times 10^3)$.

K	T	ρ	True model		Working model					
			Drop-out		GEE		WEE		CWEE	
			Type	Amount	SE	SD	SE	SD	SE	SD
50	6	0.0	MCAR	10%	494	520	489	520	586	647
			MAR-weak		492	507	487	505	594	648
			MAR-strong		489	490	486	491	616	668
		0.2	MCAR		450	469	449	472	515	555
			MAR-weak		443	474	446	478	518	571
			MAR-strong		440	466	449	482	531	594
		0.6	MCAR		317	326	328	340	364	401
			MAR-weak		318	333	333	350	366	380
			MAR-strong		315	329	344	351	373	395
100	6	0.0	MCAR	20%	427	439	424	440	442	457
			MAR-weak		421	439	419	440	440	460
			MAR-strong		413	433	416	442	452	485
		0.2	MCAR		392	400	397	407	405	413
			MAR-weak		381	403	393	417	401	421
			MAR-strong		372	397	402	437	418	444
		0.6	MCAR		283	296	310	326	293	309
			MAR-weak		275	286	313	335	292	297
			MAR-strong		265	292	339	352	303	321
200	4	0.0	MCAR	40%	357	373	355	374	360	378
			MAR-weak		345	344	345	347	352	353
			MAR-strong		334	338	346	358	358	369
		0.2	MCAR		333	341	336	346	335	345
			MAR-weak		322	342	333	355	332	351
			MAR-strong		305	316	340	339	339	336
		0.6	MCAR		248	254	273	281	249	256
			MAR-weak		235	248	275	283	243	247
			MAR-strong		224	249	299	307	250	260

Working model uses exchangeable correlation matrix.

cent of subjects have complete data. Lastly, the efficiency of CWEE was consistently poor, with the worst relative performance for the least amount of drop-out.

4. DISCUSSION

The simulation study presented in this paper illustrated that robustness to choice of working correlation in GEE does not generally hold in cases where the data are not missing completely

Table IX. Efficiency of $\hat{\beta}_4$, $K=200$, $T=4$.

ρ	True model		Working model				
	Drop-out		GEE		WEE	CWEE	
	Type	Amount	Indep	Exch	Exch	Indep	Exch
0.0	MCAR	10%	100	100	100	67	65
	MAR-weak		101	101	101	68	64
	MAR-strong		101	101	101	58	54
0.2	MCAR		99	99	99	65	70
	MAR-weak		82	101	99	56	62
	MAR-strong		38	103	101	52	57
0.6	MCAR		88	100	96	50	65
	MAR-weak		18	101	99	54	69
	MAR-strong		3	96	95	45	59
0.0	MCAR	20%	101	100	100	92	90
	MAR-weak		101	101	100	89	87
	MAR-strong		103	103	101	86	84
0.2	MCAR		96	100	98	83	90
	MAR-weak		55	101	100	83	89
	MAR-strong		14	103	100	77	85
0.6	MCAR		80	100	94	69	89
	MAR-weak		6	99	94	71	90
	MAR-strong		1	65	92	64	87
0.0	MCAR	40%	95	94	94	92	91
	MAR-weak		104	104	102	99	98
	MAR-strong		98	98	84	84	79
0.2	MCAR		97	101	98	95	100
	MAR-weak		32	99	95	92	96
	MAR-strong		8	87	94	89	95
0.6	MCAR		60	86	73	67	92
	MAR-weak		3	79	67	61	106
	MAR-strong		1	26	70	69	111

Efficiency $= 100 \times \text{mse}(\hat{\beta}_{T,\text{ML}})/\text{mse}(\hat{\beta}_{T,\text{EE}})$, where EE is the estimating equations method and ML is maximum likelihood. WEE-indep efficiency is the same as the reported WEE-exch efficiency rounded to nearest whole per cent.

at random [2]. Additionally, we demonstrated that GEE may perform poorly when data are MAR even when the correlation structure is correctly specified. Lipsitz *et al.* [10] have proposed a modified GEE approach for handling missing response data that is based upon an alternative estimator for ρ. In a simulation study of longitudinal binary data with two time points, it yielded regression parameter estimates with less bias than the standard GEE when the data are MAR and the correlation structure has been correctly specified.

Under MAR, the weighted GEE using observation specific weights [12] performed well, in terms of bias and efficiency, for estimating trends in longitudinal data. Its success, however, depended upon a correctly specified model for the missing data mechanism. Under a misspecified drop-out model, we demonstrated that it is possible to have worse performance with weighted GEE than with GEE. Lipsitz *et al.* [10] reported similar findings for a model with both an observation and a cluster-level covariate.

The results of the simulation study suggest that the weighted GEE procedure with cluster-level weights [13] should not be used as a general method. It can be considerably less efficient than the observation-weighted GEE. This finding is consistent with an earlier report in a different setting [15]. To see why the cluster-level weights may be very inefficient, reconsider the case where the missingness probability is small ($\alpha_0 = 3.0$). The probability of observing only the first response is low, and the few such clusters receive a large weight, even though these clusters contain the least information on the time trend parameter β_T. Indeed, the weighted GEE with cluster-level weights gave large bias when $K = 50$ with minimal drop-out.

We found that variance estimators for all six methods tended to underestimate the true variance of the regression parameter estimates to similar degrees. Adjustments to the weighted GEE variance estimators like those proposed for GEE [21] may be needed in small samples. There may be some interest in using (4) with $S_i = 0,\ i = 1, \ldots, K$ since these may be easily obtained by adapting existing software. This approach is conservative. In results for $K = 200$, $n = 4$ (not shown), we found that such procedures resulted in the average of the standard errors being up to 15 per cent greater than the corresponding Monte Carlo estimate.

The simulation study did not reveal an efficiency gain by using the correct exchangeable correlation in WEE relative to an independence working correlation assumption. In other settings, the choice of the working correlation matrix in equation (2) may affect efficiency [14]. For 200 clusters with maximum cluster size of 4, we found that the weighted GEE with observation level weights had high efficiency except in the notable case of severe drop-out (40 per cent conditional drop-out or 22 per cent completers) and high intracluster correlation ($\rho = 0.6$). Relative to unweighted GEE and CWEE, it was overall the most efficient. We did not examine the semi-parametric efficient estimator [22].

APPENDIX

Suppose we wish to simulate Y, a T-vector of Bernoulli variates with mean vector π and covariance matrix V. For $t = 2, \ldots, T$, define $Z_t = (Y_1, \ldots, Y_{t-1})^\top$, $\mu_t = E(Z_t)$, $G_t = \mathrm{cov}(Z_t)$, and $s_t = \mathrm{cov}(Z_t, Y_t)$. Note that G_t and s_t are determined from V. For a given (π, V), a $(t-1)$-vector b_t is defined as $b_t = G_t^{-1} s_t$ $(t = 2, \ldots, T)$. The conditional linear family (introduced in an unpublished work by Bahjat Qaqish) is defined by

$$\begin{aligned} v_t = v_t(z_t; \pi, V) &:= P(Y_t = 1 \mid Z_t = z_t) = \pi_t + b_t^\top (z_t - \mu_t) \\ &= \pi_t + \sum_{j=1}^{t-1} b_{tj}(y_j - \pi_j) \quad (t = 2, \ldots, T) \end{aligned} \tag{A1}$$

The simulation algorithm proceeds as follows. First, simulate Y_1 as Bernoulli with mean π_1, then for $t = 2, \ldots, T$, simulate Y_t as Bernoulli with conditional mean v_t given by (A1). It then follows that $E(Y)=\pi$ and for $1<t\leqslant T$, $\text{cov}(Z_t, Y_t)=\text{cov}(Z_t, b_t^{\top} Z_t)=G_t b_t=s_t$. The vector Y thus obtained has the required mean, π, and covariance, V. There are some restrictions on allowable π and V as discussed by Qaqish.

The full joint distribution of Y, whose explicit specification is not required, can be computed via (A1). For any valid (π, V) that is reproducible by the conditional linear family, there is a corresponding unique value of the $2^T \times 1$ vector θ of joint probabilities. For $T=4$ and exchangeable correlation, let $\theta_{ijkl}=P(Y_1=i, Y_2=j, Y_3=k, Y_4=l)$ and $\theta^{\mathrm{T}}(\pi,\rho)=\{\theta_{0000}, \theta_{0001}, \ldots, \theta_{1111}\}$ be the 2^4 vector of joint probabilities with the later indices changing fastest. Letting $\pi_0=(0.332, 0.348, 0.362, 0.378)$, the simulation results in Tables III, IX and the bottom thirds of Tables VII and VIII were based upon generating $T=4$ correlated Bernoulli variates from:

$$(0.173\ 0.105\ 0.098\ 0.060\ 0.092\ 0.056\ 0.052\ 0.032\ 0.086\ 0.052\ 0.049\ 0.030\ 0.046\ 0.028\ 0.026\ 0.016)$$

$$(0.280\ 0.082\ 0.075\ 0.044\ 0.069\ 0.041\ 0.038\ 0.040\ 0.063\ 0.037\ 0.035\ 0.037\ 0.032\ 0.034\ 0.032\ 0.062)$$

$$(0.466\ 0.045\ 0.038\ 0.022\ 0.032\ 0.019\ 0.017\ 0.030\ 0.027\ 0.016\ 0.014\ 0.026\ 0.012\ 0.022\ 0.017\ 0.199)$$

the values of $\theta^{\mathrm{T}}(\pi_0, 0), \theta^{\mathrm{T}}(\pi_0, 0.2)$, and $\theta^{\mathrm{T}}(\pi_0, 0.6)$, respectively.

ACKNOWLEDGEMENTS

We thank Bahjat F. Qaqish for help in regards to generating correlated binary data, and Michael E. Miller and two anonymous referees for comments leading to improvements in the paper. An earlier draft was written while J. Preisser was on the faculty of Wake Forest University School of Medicine, Winston-Salem, NC. The work was partially supported by grant IRG-198A from the American Cancer Society. K. Lohman acknowledges support from grant AG14131 from the National Institutes of Health.

REFERENCES

1. Diggle PJ, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data*. Oxford University Press: Oxford, 1994.
2. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
3. Hendricks SA, Wassell JT, Collins JW, Sedlak SL. Power determinations for geographically clustered data using generalized estimating equations. *Statistics in Medicine* 1996; **15**:1951–1960.
4. Carr GJ, Portier CJ. An evaluation of some methods for fitting dose–response models to quantal–response developmental toxicology data. *Biometrics* 1993; **49**:779–791.
5. Bieler GS, Williams RL. Cluster sampling techniques in quantal response teratology and developmental toxicity studies. *Biometrics* 1995; **51**:764–776.
6. Sharples K, Breslow N. Regression analysis of correlated binary data: some small sample results for the estimating equation approach. *Journal of Statistical Computation and Simulation* 1992; **42**:1–20.
7. Emrich LJ, Piedmonte MR. On some small sample properties of generalized estimating equation estimates for multivariate dichotomous outcomes. *Journal of Statistical Computation and Simulation* 1992; **41**:19–29.
8. Lipsitz SR, Fitzmaurice GM, Orav EJ, Laird NM. Performance of generalized estimating equations in practical situations. *Biometrics* 1994; **50**:270–278.
9. Gunsolley JC, Getchell C, Chinchilli VM. Small sample characteristics of generalized estimating equations. *Communications in Statistics-Simulations* 1995; **24**:869–878.
10. Lipsitz SR, Molenberghs G, Fitzmaurice GM, Ibrahim J. GEE with Gaussian estimation of the correlations when data are incomplete. *Biometrics* 2000; **56**:528–536.
11. Preisser JS, Galecki AT, Lohman KK, Wagenknecht LE. Analysis of smoking trends with incomplete longitudinal binary responses. *Journal of the American Statistical Association* 2000; **95**:1021–1031.
12. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995; **90**:106–121.
13. Fitzmaurice GM, Molenberghs G, Lipsitz SR. Regression models for longitudinal binary responses with informative drop-outs. *Journal of the Royal Statistical Society*, *Series B* 1995; **57**:691–704.

14. Troxel AB. A comparative analysis of quality of life data from a southwest oncology group randomized trial of advanced colorectal cancer. *Statistics in Medicine* 1998; **17**:767–779.
15. O'Hara-Hines RJ, Hines WGS, Friesen TG. A comparison of two drop-out weighting schemes in the analysis of clustered data with categorical and continuous responses. *Journal of Agricultural*, *Biological*, *and Environmental Statistics* 1999; **4**:203–216.
16. Miller ME, Ten Have TR, Reboussin BA, Lohman KK, Rejeski WJ. A marginal model for analyzing discrete outcomes from longitudinal surveys with outcomes subject to multiple cause non-response. *Journal of the American Statistical Association* 2001; **96**:844–857.
17. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Wiley: New York, 1987.
18. Troxel AB, Lipsitz SR, Brennan TA. Weighted estimating equations with nonignorably missing response data. *Biometrics* 1997; **53**:857–869.
19. Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* 1998; **93**:1321–1339.
20. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd edn. Chapman and Hall: London, 1989.
21. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; **57**:126–134.
22. Robins JM, Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* 1995; **90**:106–121.

